A Combined Approach to Subsurface Scattering-Draft

Adrian Mayorga
Supervised by Dr. Fussell
University of Texas at Austin
Department of Computer Sciences

April, 2008

Abstract

Current techniques for rendering multilayered translucent materials such as skin, marble, jade, and paper can achieve startling realism by modeling subsurface scattering. However, due to performance concerns, all of these techniques require fixing the material properties, including the depth and scattering parameters of each layer, before rendering begins. This makes it impractical to calculate scattering profiles on a per pixel basis to render objects that have spatially varying parameters. We introduce a new method that combines ideas from two techniques to enable the rendering of multilayered materials with spatially varying parameters. Specifically, we combine hierarchical integration with a sum of Gaussians approximation of subsurface scattering diffusion profiles. This speeds up the process of computing the interactions between individual layers by orders of magnitude, allowing parameters to vary across the surface. We show that this technique allows for rendering of translucent materials with spatially varying properties with relatively low overhead.

1 Introduction

Many of the materials that we encounter in everyday life, like paper, skin, marble, jade, and milk, are translucent. Translucency is actually due to a phenomenon known as subsurface scattering.

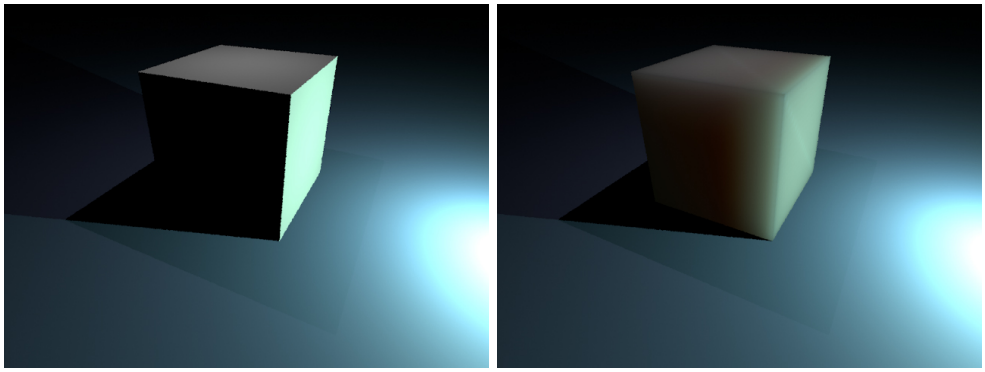
When a photon hits the surface of an object, instead of being immediately reflected or internally absorbed as in opaque materials, in translucent materials the photons penetrate the surface. Once inside the surface, the photons scatter on internal structures before exiting through the surface at another location, creating a soft appearance. Simulating subsurface scattering in software renderers drastically increases the realism of translucent objects.

There have been several methods for computing this simulation. Traditionally subsurface scattering was largely ignored and simply approximated as Lambertian diffuse reflection. Hanrahan and Krueger (1993) introduced an analytic solution for single scattering. Dorsey et al. (1999) treated the material as a participating medium and used Photon Map tracing. Pharr and Hanrahan (2000) introduced a method that uses Monte Carlo integration and scattering equations to solve the radiative transfer equation. These two methods are capable of simulating full subsurface scattering, but rendering with them takes a very long time.

Jensen et al. (2001) introduces a dipole-based method that simulates diffuse scattering. In Jensen and Buhler (2002) this computation is sped up by decoupling it into two parts, and in Donner and Jensen (2005) a multipole method is introduced to allow rendering of thin slabs and multilayered materials such as skin.

Other work has also been done in image-space subsurface scattering, which approximates diffuse scattering on the surface through specific texture blurring. d'Eon et al. (2007) introduce a texture-based approximation of Donner and Jensen (2005) based on a sum of Gaussians fit to their model.

Our method combines two of the previously mentioned methods. Specifically, we combine Jensen and Buhler (2002) decoupling with d'Eon et al. (2007) sum of Gaussian fit. This enables us to render materials which have varying spatial properties, such as varying slab depth through the surface of the object.



(a) No subsurface scattering (b) With subsurface scattering

Figure 1: This figure illustrates the effect that adding a subsurface scattering simulation can have.

2 Related Work

The method of Dorsey et al. (1999) relies on tracing many photons (in a similar manner to photon mapping (Jensen (1996)) through the inside of the surface and modeling each scattering event. This works well in practice for low scattering materials, but the large number of photons needed to correctly simulate the appearance of highly scattering materials, such as skin or milk, makes this method impractical, although with enough time it can produce "ground truth" renderings.

To address this problem Jensen et al. (2001) introduced a model based on a twin light source (dipole) approximation. This model assumes that the light scatters inside of the material in an isotropic (same in all directions) manner, which works well for materials in which photons scatter several times before exiting. Because of this assumption, it is possible to model the way light scatters as a 1D function dependent on the distance to a light sample. Jensen calls this function a diffusion profile, and this is what the dipole approximates.

Jensen and Buhler (2002) introduced an optimization to this method that decouples the computation into two phases. The first phase calculates the irradiance on the surface of the material and stores it in a hierarchical structure (kd-tree, octree) for later access. It is sufficient to store only the irradiance since the light is assumed to scatter in an isotropic manner, and the angle at which light arrives loses any significance. The second phase uses the computed irradiance samples to hierarchically integrate the scattering model.

Later Donner and Jensen (2005) introduced an extension to his dipole model that allows the correct rendering of thin slabs of material, such as a piece of paper, and multilayered materials like skin. The multipole is basically a sum of dipoles, each placed at increasingly distant locations. To render a multilayered material, Jensen recursively combines the diffusion profiles of two layers by convolving their response.

This process involves several FFT, 2D texture additions and reverse FFT, which makes it impractical to perform on a per pixel basis. For this reason, is it necessary to fix all material parameters, including slab depths in a multilayered material, at pre-computation time and store the results in a table. It should be noted that the multipole model does have some caveats. The multipole assumes that each layer of the material scatters light in a diffuse manner, which means that the depth of each slab must be at least a few mean free paths. Furthermore, to decouple each layer, it is assumed that the interactions between each of the layers are due to multiple scattering only. This allows Donner and Jensen (2005) to simulate each layer separately and compute the combined result by convolving the profiles of each layer. While this is not entirely physically correct, the approximation is accurate close to the source, where accuracy is most important to the overall appearance. Donner and Jensen (2005) also show that the difference in total reflectance and transmittance between the multipole model and a Monte Carlo simulation is minimal.

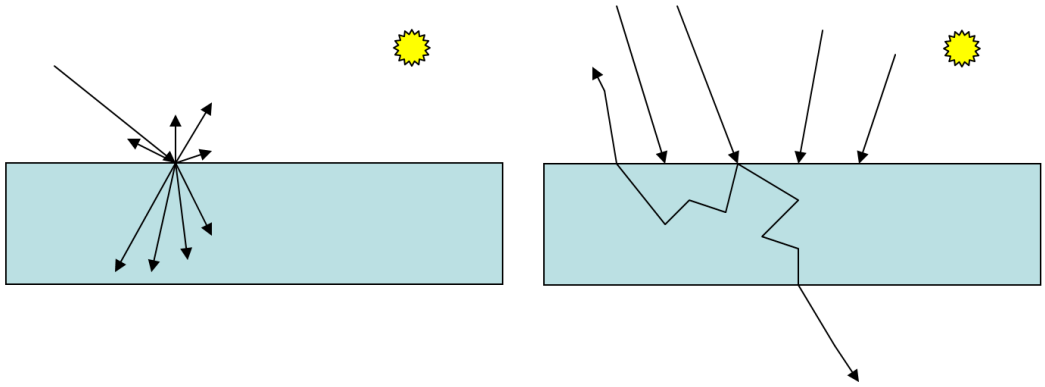
Other approaches have used texture-space diffusion. To simulate diffusion, the incoming light is stored into an off-screen texture. Then, to approximate the scattering of light, the light texture is

blurred and applied to the model. This effectively only simulates scattering through the surface and ignores global effects, such as when light bleeds from one side of an object to the other. These types of techniques are not generally physically based and usually used where performance is more important that accuracy.

d'Eon et al. (2007) introduce a texture space approximation to the dipole/multipole model. They approximate the diffusion profiles predicted by the multipole with a sum of Gaussians. They then leverage certain properties of Gaussians to convolve the off-screen irradiance texture efficiently. In addition to convolving the irradiance, they use a stretch-correction technique that diminishes the artifacts created by texture stretching. They also use an extension to Translucent Shadow Maps(Dachsbacher and Stamminger (2003)) to capture global effects, such as light bleeding through the ear when a character is lit from behind.

3 Background

3.1 Subsurface Scattering



(a) BRDF/BTDF (b) BSSRDF

Figure 2: Comparison of BRDF/BTDF and BSSRDF

Normal shading models typically use a Bidirectional Reflectance Distribution Function (BRDF) to simulate opaque material properties. The BRDF relates the amount of light that is reflected in a specific direction w_o from incident light coming from direction w_i. A similar family of functions, called Bidirectional Transmittance Distribution Functions (BTDF), models the way that light is transmitted through transparent materials (such as glass).

By combining a BTDF and BRDF for a single material, complex light-material interactions such as light bouncing off brushed metal and specular reflections, or light diffracting through glass can be faithfully reproduced. However, since the BRDF and the BTDF assume that light is either reflected, absorbed, or transmitted at a single point it cannot accurately simulate the complex interactions of subsurface scattering found in translucent materials.

This is because the photons that hit the surface of a translucent material actually penetrate and are scattered in several directions by internal structures before they are transmitted or reflected back through some other point. This phenomenon can be described by a Bidirectional Scattering Surface Reflectance Distribution Function (BSSRDF) (Nicodemus et al. (1992)). A BSSRDF relates the amount of light being reflected at a point in a particular direction as a function of all of the incoming light throughout the entire surface. Figure 2 illustrates the differences between a BRDF a BSSRDF

In fact, the BRDF is only an approximation of the BSSRDF. With this in mind the classical reflection equation

$$L_o(x, \vec{\omega_o}) = \int_{2\pi} f(x, \vec{\omega_o}, \vec{\omega_i}) L_i(x, \vec{\omega_i})(\vec{n} \cdot \vec{\omega_i}) d\omega_i$$

Turns into

$$L_o(x_o, \vec{\omega_o}) = \int_A \int_{2\pi} S(x_i, \vec{\omega_i}, x_o, \vec{\omega_o}) L_i(x, \vec{\omega_i})(\vec{n} \cdot \vec{\omega_i}) d\omega_i$$

Where $L_i(x_i, \omega_i)$ is the incoming radiance at point x_i in direction ω_i, L_o is the outgoing radiance, f is the BRDF and S is the BSSRDF.

One approach to simulating this is photon tracing (Dorsey et al. (1999)). This method works especially well for materials that do not exhibit a high degree of scattering, but as scattering increases, the number of photons, and the length of the path for each photon needed to faithfully reproduce a material's appearance increases rapidly. For such materials scattering can generally be separated into single scattering (one scattering event) and multiple scattering. As the scattering character of a material increases, so does the importance of multiple scattering relative to single scattering. In highly scattering materials multiple scattering dominates the appearance.

3.2 Diffuse Scattering

Inside highly scattering homogeneous materials, the scattering of photons quickly becomes isotropic. This allows for subsurface scattering to be approximated by diffuse reflection. Assuming that the geometry is locally planar, the scattering diffuse reflectance of an object can be simulated by using a 1D function that relates the amount of light reflected back to the surface to the distance from the incoming beam of light. This 1D function is called a diffusion profile and can be measured by shining a laser beam perpendicular to a block of material and capturing the spreading of light around the point. Figure 3 is an example of such a profile. However, as Jensen et al. (2001) show, it can also be analytically approximated faithfully. Also note that the BSSRDF can now be approximated by:

$$S(x_i, \vec{\omega_i}, x_o, \vec{\omega_o}) = \frac{1}{\pi} F_t(x_i, \vec{\omega_i}) R(|x_i - x_o|) F_t(x_o, \vec{\omega_o})$$

where R is the diffusion profile.

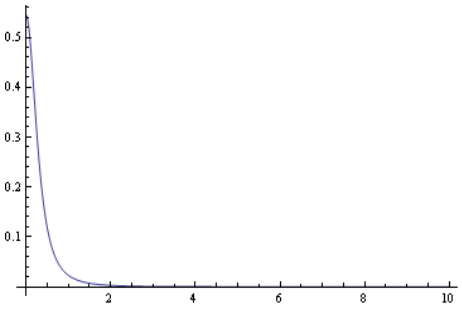


Figure 3: Diffusion profile for green light in marble

3.2.1 Dipole Approximation

The scattering of photons inside the material can be modeled by treating the object as a participating medium with the equation of radiative transfer (Chandrasekhar (1960))

$$(\vec{\omega} \cdot \nabla) L(x, \vec{\omega}) = -\sigma_t L(x, \vec{\omega}) + \sigma_s L_i(x, \vec{\omega}) + s(x, \vec{\omega}), \tag{1}$$

where L is the radiance, s is a source term, σ_s is the scattering coefficient, σ_a is the absorption coefficient, $\sigma_t = \sigma_a + \sigma_s$ and Li is the in-scattered radiance:

$$L_i(x, \vec{\omega}) = \int_{4\pi} p(\vec{\omega}, \vec{\omega}') L(\vec{\omega}, \vec{\omega}') d\vec{\omega}'.$$

Here p is a phase function which describes the distribution of the scattering. Solving equation 1 even for isotropic scattering is a difficult task. To get around this Jensen et al. use a spherical harmonic series to approximate the radiance:

$$L(x,\vec{\omega}) = \frac{1}{4\pi}\phi(x) + \frac{3}{4\pi}\vec{\omega}\cdot\vec{E}(x),$$

where ϕ is the incoming radiant flux.

By substituting this equation into equation (1) and integrating, Jensen et al. (2001) eventually arrive at the classical diffusion equation, which has a simple solution for an isotropic light source in an infinite medium:

$$\phi(x) = \frac{\Phi}{4\pi D}\frac{e^{-\sigma_{tr}r(x)}}{r(x)}$$

where Φ is the power of the light source, r is the distance to the source, D is the diffusion constant $D = \frac{1}{3\sigma_t'}$, $\sigma_t' = \sigma_a + \sigma_s'$ and $\sigma_{tr} = \sqrt{3\sigma_a\sigma_t'}$.

However, since an object has a surface, this introduces a boundary condition which must be satisfied. This boundary condition is that the net inwards diffuse radiance is zero at the surface. In other words, the diffuse radiance into the surface is equal to the diffuse radiance internally reflected outward by the material:

$$\int_{2\pi_-} L(x,\vec{\omega})(\vec{\omega}\cdot\vec{n}_-)d\omega = F_{dr}\int_{2\pi_+} L(x,\vec{\omega})(\vec{\omega}\cdot\vec{n}_+)d\omega \tag{2}$$

at $z = 0$, where F_{dr} is the Fresnel diffuse reflectance approximated according to Egan et al. (1973) by:

$$F_{dr} \approx \begin{cases} -.4399 + \frac{.7099}{\eta} - \frac{.3319}{\eta^2} + \frac{.0636}{\eta^3}, & \eta < 1 \\ -\frac{1.4399}{\eta^2} + \frac{.7099}{\eta} + .6681 + .0636\eta, & \eta > 1 \end{cases}$$

where η is the index of refraction.

Substituting this boundary condition into equation (2) gives:

$$\phi(r) - 2AD\frac{\partial\phi(r)}{\partial z} = 0 \text{ at } z = 0$$

where $A = (1 + F_{dr})/(1 - F_{dr})$.

This means that the net flux will vanish at a distance $z_b = 2AD$ above the surface. Since the scattering of incoming light is approximated by placing a light source at a depth of one mean free path, $l = 1/\sigma_t'$ beneath the surface, the boundary condition can be satisfied by placing a negative light source at $2z_b + l$ above the surface. This results in the dipole model presented by Jensen(2001)

$$R(r) = \frac{\alpha' z_r(1+\sigma_{tr}d_r)e^{-\sigma_{tr}d_r}}{4\pi d_r^3} - \frac{\alpha' z_v(1+\sigma_{tr}d_v)e^{-\sigma_{tr}d_v}}{4\pi d_v^3}$$

where $\alpha' = \sigma_s'/\sigma_t'$ is the reduced albedo, $z_r = l = 1/\sigma_t$ and $z_v = -l - 2z_b$ are the positions of the real and virtual light sources and where $z = 0$ is the surface of the object, $d_r = \sqrt{r^2 + z_r^2}$ and $d_v = \sqrt{r^2 + z_v^2}$ are the distances to the real and virtual light. Here $r = |x_o - x_i|$ is the distance.

3.2.2 Multipole Approximation

The dipole model has the limitation that it can only simulate homogeneous materials in objects that are optically thick. This means that in the case of thin geometry, the dipole will overestimate the amount of light that is reflected back to the surface and underestimate the amount of light that is transmitted through the object. To fix this Donner and Jensen (2005) consider a thin slab of material of depth d, introducing a second boundary condition, the same as equation (2) but at $z = d$, which gives:

$$\phi(r) - 2AD\frac{\partial\phi(r)}{\partial z} = 0 \text{ at } z = d.$$

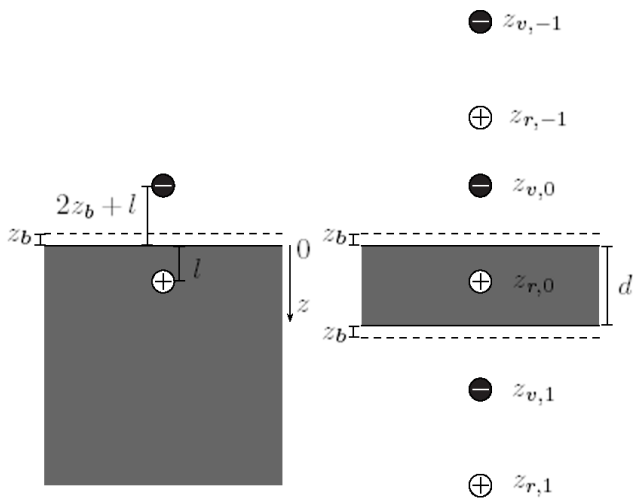


Figure 4: Comparison of the dipole model(left) and the multipole model(right)

This condition can be satisfied by reflecting the original dipole over a line at $d + z_b$. However, adding this dipole pair invalidates the original boundary condition but can be corrected by reflecting this new dipole about z_b. In order to satisfy both boundary equations at once, an infinite sum of dipoles above and below the original one is needed. The sources of a particular dipole pair are then given by:

$$z_r = 2i(d + 2z_b) + l$$
$$z_v = 2i(d + 2z_b) - l - 2z_b, i = -n, ..., n$$

where $2n + 1$ is the number of dipole pairs.

Therefore, the reflectance diffusion profile for a thin slab is:

$$R(r) = \sum_{i=-n}^{n} \frac{\alpha' z_{r,i}(1 + \sigma_{tr} d_{r,i}) e^{-\sigma_{tr} d_{r,i}}}{4\pi d_{r,i}^3} - \frac{\alpha' z_{v,i}(1 + \sigma_{tr} d_{v,i}) e^{-\sigma_{tr} d_{v,i}}}{4\pi d_{v,i}^3}.$$

The previous derivation assumes that the indices of refraction are the same above and below the slab, which may not always be the case. Additionally, light is now transmitted through the slab. This light can be modeled in much the same way that the reflection was modeled, but adjusting for the depth of the slab. To do this we have to mirror all of the dipoles about the lower edge at $z = d$. Since the indices of refraction may not be the same, each slab now has 4 diffusion profiles, R^+, R^-, T^+, T^- (see Figure 5). Here the $+$ and $-$ denote forward or backwards direction. This gives us the following equations for the profiles:

$$
\begin{aligned}
R^+(r) =& M_P(r) \text{ for} \\
& z_r = 2i(d + z_b(0) + z_d(d)) + l \\
& z_v = 2i(d + z_b(0) + z_d(d)) - l - 2z_b(0) \\
R^-(r) =& M_P(r) \text{ for} \\
& z_r = 2i(d + z_b(d) + z_d(0)) + l \\
& z_v = 2i(d + z_b(d) + z_d(0)) - l - 2z_b(d) \\
T^+(r) =& M_P(r) \text{ for} \\
& z_r = d - \{2i(d + z_b(0) + z_d(d)) + l\} \\
& z_v = d - \{2i(d + z_b(0) + z_d(d)) - l - 2z_b(0)\} \\
T^-(r) =& M_P(r) \text{ for} \\
& z_r = d - \{2i(d + z_b(d) + z_d(0)) + l\} \\
& z_v = d - \{2i(d + z_b(d) + z_d(0)) - l - 2z_b(d)\}
\end{aligned}
$$

where $M_P(r)$ is a general multipole:

$$
M_P(r) = \sum_{i=-n}^{n} \frac{\alpha' z_{r,i}(1 + \sigma_{tr} d_{r,i}) e^{-\sigma_{tr} d_{r,i}}}{4\pi d_{r,i}^3} - \frac{\alpha' z_{v,i}(1 + \sigma_{tr} d_{v,i}) e^{-\sigma_{tr} d_{v,i}}}{4\pi d_{v,i}^3}
$$

$z_b(0)$ and $z_b(d)$ are calculated by using $A(0)$ and $A(d)$ using the relative indices of refraction at the top and bottom boundaries. Note that if the indices of refraction are the same at the top and bottom, then $z_b(0) = z_b(d)$ and $R^+(r) = R^-(r), T^+(r) = T^-(r)$. Also note that the equations given for T^+ and T^- are not the same as presented in Donner and Jensen (2005) since there was a slight error there. They have been corrected here.

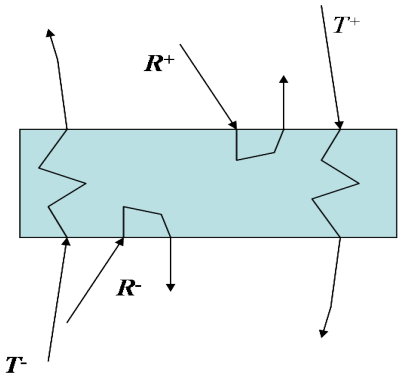


Figure 5: The four different profiles generated by the multipole model

3.2.3 Multilayered Materials

The multipole can be used to render multilayered materials. Each layer is represented as a thin homogeneous slab, which are then stacked on top of each other. To account for the interaction between layers, a convolution can be used to combine the responses of two layers. For example, to calculate the combined forward Transmittance T_{12}^+ of a two-layered material with transmittance profiles T_1^+ and T_2^+, we convolve then together to get $T_{12}^+ = T_1^+ * T_2^+$.

This assumes that light transmits through the first layer and then through the second layer. However, once light scatters through the first layer, it can be reflected back by the second layer, then back to the second, and finally transmitted through the second. The mutual reflection between the interface of the first two layers can occur an infinite number of times, and must be accounted for, thus

$$\begin{aligned} T_{12}^{+} =& T_1^{+} * T_2^{+} + T_1^{+} * R_2^{+} * R_1^{-} * T_2^{+} + \\ & T_1^{+} * R_2^{+} * R_1^{-} * R_2^{+} * R_1^{-} * T_2^{+} + ... \end{aligned}$$

This series can be transformed into frequency space where the convolutions turn into multiplication

$$\begin{aligned} T_{12}^{+} =& T_1^{+} T_2^{+} + T_1^{+} R_2^{+} R_1^{-} T_2^{+} + \\ & T_1^{+} R_2^{+} R_1^{-} R_2^{+} R_1^{-} T_2^{+} + ... \\ =& T_1^{+} T_2^{+} \left(\sum_{i=0}^{\infty} (R_2^{+} R_1^{-})^i \right) \end{aligned}$$

Assuming that $R_2^{+} R_1^{-} < 1$ the equation can be simplified to

$$T_{12}^{+} = \frac{T_1^{+} T_2^{+}}{1 - R_2^{+} R_1^{-}}$$

The same derivation can be applied to obtain similar equations for T_{12}^{-}, R_{12}^{+},and R_{12}^{-}

$$R_{12}^{+} = R_1^{+} + \frac{T_1^{+} R_2^{+} T_1^{-}}{1 - R_2^{+} R_1^{-}}$$

For the analogous backwards scattering simply flip all of the directions.

To obtain the combined profiles of a two layered material, first a 2D radial Fourier transform, or a 1D Hankel transform, is applied to each of the 4 profiles for each slab. Then the above formulas are used to solve the geometric series for each profile,and finally an inverse Fourier or Hankel transform is applied. If the material has more than 2 layers, the entire process is repeated recursively.

3.3 Sum of Gaussians Approximation

d'Eon et al. (2007) present a method for realistic real-time rendering of human skin. They achieve this by fitting a sum of Gaussians to the profiles predicted by the dipole/multipole method of Jensen et al. The specific Gaussian is given by

$$G(v, r) = \frac{1}{2\pi} e^{\frac{-r^2}{2v}} \tag{3}$$

where v is the variance. These Gaussians where chosen because their 2D radial response is equal to 1

$$\int_0^{\infty} 2\pi r G(v, r) dr = 1.$$

d'Eon et al. (2007) provide two distinct methods for fitting a sum of Gaussians to a diffusion profile. The first method fits a set number of Gaussian terms to a discretized diffusion profile using optimization code. The second method analytically approximates each pole in the multipole model with 4 Gaussians to obtain quick analytic fit at the expense of more Gaussian terms. In our work we only use the second method.

For the first method, a sum of k Gaussians can be fit to a particular profile $M_P(r)$ by minimizing

$$\int_0^{\infty} r * \left(M_P(r) - \sum_{i=1}^{k} w_i G(v_i, r) \right)^2 dr$$

where w_i and v_i are allowed to vary. With this method, 4-8 Gaussians are usually enough to approximate a multipole with little error, measured by

$$\frac{\sqrt{\int_0^\infty r\left(M_P(r)-\sum_{i=1}^k w_i G(v_i,r)\right)^2 dr}}{\sqrt{\int_0^\infty r\left(M_P(r)\right)^2 dr}}$$

For the second method, the authors provide a way to empirically fit a sum of Gaussians to any dipole/multipole. Each dipole/multipole is really just a sum of several poles, each of the same general form

$$P(r,\sigma_{tr},z)=\frac{(1+\sqrt{z^2+r^2}\sigma_{tr})e^{-\sqrt{z^2+r^2}\sigma_{tr}}}{(z^2+r^2)^{\frac{3}{2}}}$$

which has power

$$P_d(\sigma_{tr},z)=\int_0^\infty P(r,\sigma_{tr},z)rdr=\frac{e^{-\sigma_{tr}|z|}}{|z|}$$

Any pole equation P can then be fit by using 4 Gaussians of variances

$$\{0.2304v_0, 0.7225v_0, 2.6569v_0, 13.6v_0\}$$

and weights

$$w_i = C_i(|\sigma_{tr}z|)P_d(\sigma_{tr},z)$$

where C_i is one of the following

$$C_1(x) = -0.3454 + \frac{1.2422}{x} - 1.2422\frac{e^{-0.7452x}}{x}$$
$$C_2(x) = 5.34071\left(\frac{(1-e^{-0.716x})(0.9029x-0.7401)}{x} + 0.90295e^{-0.7165x}\right)$$
$$C_3(x) = 10.2416\left(\frac{(1-e^{-1.0961x})(0.242x+0.1804)}{x} - 0.00244e^{-1.0961x}(0.2041x+0.1752)\right)$$
$$C_4(x) = 23.185\left(\frac{(1-e^{-1.433x})(0.0505-0.0395x)}{x} + 0.04425e^{-1.433x}(0.09187x+0.06034)\right)$$

This gives a a fast analytic fit to any dipole with only 8 Gaussians. For a multipole, since the contribution of each dipole pair decreases with distance, only a small number of dipole pairs are needed. We found that using 2 dipole pairs above and below the original dipole gives almost indistinguishable result from using 500. This gives a fast analytic fit to any multipole with 40 Gaussians.

Once a profile is represented as a sum of Gaussians, the convolutions needed to compute the combined response of two slabs can be done much faster. This is because the convolution of two Gaussians is another Gaussian

$$G(v_1) * G(v_2) = G(v_1+v_2)$$

which means that the convolution of two sum-of-Gaussian profiles is

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(w_i w_j G(v_i+v_j,r)). \tag{4}$$

This expressions results in a new profile of length mn which can lead to a rapid explosion in the number of terms after just calculating the first two terms a combined profile (starting with $m=40, n=40$, calculating the first two terms of T_{12} yields 64040 terms!). Even so, if the Gaussians are restricted to the set

$$\{G(v), G(v)*G(v), G(v)*G(v)*G(v), ...\} =$$
$$\{G(v), G(2v), G(3v), ...\}$$

then computing (4) is analogous to polynomial multiplication, which leads to $m+n$ terms.

4 Combining the Two Techniques

We combine the Jensen and Buhler (2002) decoupling of irradiance and hierarchical integration technique with d'Eon's sum-of-Gaussians diffusion profile. This speeds up the computation for calculating the combined response of two slabs, enabling us to change slab parameters using a texture map. This leads to increased realism and higher quality images with relatively low overhead. Our system is implemented as an extension to PBRT (Pharr and Humphreys (2004)).

4.1 Sampling the Irradiance

The first step is to sample the irradiance on the surface and store it in a hierarchical structure for future integration. Jensen and Buhler (2002) use a point repulsion algorithm to obtain a uniform sampling over the mesh. Instead of using this, we opt for a simpler approach that still generates good results.

We perform uniform sampling on a per triangle basis. This has the advantage that it is simpler to code, and since no knowledge of other triangles is necessary, PBRT does not need to be modified extensively. Because we will store the area associated with each sample, uniform sampling is not really necessary. The only major requirement is that the sampling density be high enough to capture the effects of diffuse scattering. This can be guaranteed by ensuring that on any point of the triangle, the closest sample point is at a distance which is at most l, the mean-free path. To guarantee this, we perform our sampling in the following manner.

First consider the centroid of a triangle. The farthest point from the centroid inside of the triangle must be one of the three vertices. Let the distance to this vertex be d_v. Then, conversely, any point in the triangle must be at most d_v away from the centroid. Also note that for any triangle the centroid of that triangle divides the medians into $2:1$ ratios (see Figure 6).

Second, any triangle may be tessellated into smaller similar triangles, some of which are flipped relative to the larger triangle. This can be done by dividing each side into n equal parts and then tracing lines parallel to the other two edges at each division node (see Figure 6).

Third, all of the smaller triangles can be considered to be copies of the large triangle scaled by $\frac{1}{n}$. This means that for any point inside a small triangle, it is at most $\frac{d_v}{n}$ away from the small triangle's centroid. Now consider only those triangles that have been flipped. Each of these triangles is always surrounded by three triangles that have not been flipped. Since the centroid of a triangle divides the medians in a $2:1$ ratio, the maximum distance from the centroid of the flipped triangle to the centroid of any of the surrounding triangles is $\frac{d_v}{n}$. This means that any point inside a flipped triangle must be at most $\frac{d_v}{n}$ away from a centroid of a non-flipped triangle, which means that any point on the *large triangle* is at most $\frac{d_v}{n}$ away from the centroid of a non-flipped triangle.

Given the mean free path l, we can easily find integer n such that $\frac{d_v}{n} \leq l$. We then use this n to divide the triangle and generate a uniform sampling pattern.

Once given the sample points, the irradiance, area, texture coordinates, normal, and position is calculated and stored in a sample structure. The samples are then added to an octree. As the octree is constructed, each node of the octree stores an average sample, which summarizes all of its children. The average irradiance is calculated by taking the average of the children's irradiance weighted by area. The average position is taken as the average weighted by the irradiance of the children.

4.2 Calculating Diffusion Profiles

To calculate the diffusion profiles for a given material we use a small variation of the analytic fit in section 3.3. Instead of directly using the 40 Gaussians of the fit, we map them to the set $\{G(v), G(2v), G(3v), ...\}$ with $v = 0.05$ and 50 terms. To do this we note that any Gaussian can be approximated by the linear combination of two other Gaussians.

$$w_{orig}G(v_{orig}) \approx w_aG(v_a) + w_bG(v_b)$$

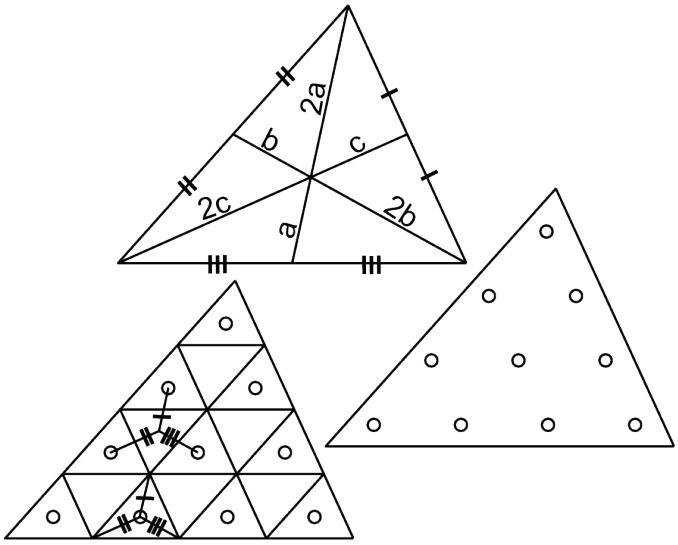


Figure 6: Top: centroid divides medians in a 2 : 1 ratio. Left: tessellated triangle, where $n = 4$. Note that samples are not needed on all of the triangles. Right: final sample positions for this triangle

where v_a and v_b are the two closest variances in our set to v_{orig} (v_b is the greater of the two) and the weights equal

$$w_a = w_{orig} \frac{v_b - v_{orig}}{v_b - v_a}$$
$$w_b = w_{orig} \frac{v_{orig} - v_a}{v_b - v_a}$$

Notice that as v_{orig} gets closer to either v_a or v_b, w_a or w_b becomes w_{orig} and the other 0. Also note that $w_a + w_b = w_{orig}$ in all cases. This means that even if the approximation is not good, the total diffuse response will not be changed, just the shape, thus conserving energy. There is one special case however. If the variance v_{orig} is larger than v_{50} then we simply accumulate the the weight w_{orig} in w_{50}. We can do this since as the variances get larger, the weights, and therefore their contribution to the entire profile, tend to get smaller. Doing this instead of trying to approximate a Gaussain of very large variance with the Gaussians $G(v_{49}), G(v_{50})$ is better since it prevents overestimation at larger radii.

When performing the convolutions of two profiles, we fix the maximum length of the resulting sum to 100 terms. The weight of any terms that would come after this is accumulated in w_{100}.

In order to facilitate fast evaluation, we remap the resulting profile after convolutions to the set

$$\{G(v), G(2v), G(3v), G(4v), G(1),$$
$$G(2), G(5), G(10), G(50)\}.$$

This speeds up evaluation since we only need to evaluate 9 terms as opposed to 100.

When the samples are being created, a texture lookup is performed to retrieve the local scattering parameters. Then the above computation is performed and the sparse diffusion profiles are stored inside of the sample structure. During rendering, the system will integrate the per-sample profiles to get the overall picture. This assumes that each sample point has a different, locally homogeneous neighborhood.

4.3 Integrating the Diffusion

We use the same integration that Jensen and Buhler (2002) uses. To calculate the diffuse response at a point x, we traverse the octree from the top down, starting from the root. At each level of the tree, we calculate whether the node is small enough based on an approximate solid angle

$$\Delta\omega = \frac{A_n}{\|x - P_n\|^2}.$$

If $\Delta\omega < \epsilon$ then we can use the node directly, otherwise we pass x to all of the children. If the node contains x we always descend. In the event that we reach a leaf node, the contribution of the node is the sum of all of the sample points contained by the node.

The contribution of a sample point (or an average node) for a point x is computed by

$$M_{o,p}(x) = F_{dt}(x)D_{iff}(r)E_pA_p$$

where $F_{dt} = 1 - F_{dr}$, $r = |x - P|$ is the distance from the sample point to the exiting point, E_p is the irradiance at sample P, A_p is the area, and $D_{iff}(r)$ combines the reflectance and transmittance profiles

$$D_{iff}(r) = \frac{1}{2}(\vec{n}_x \cdot \vec{n}_p + 1)R^+(r) + \frac{1}{2}(1 - \vec{n}_x \cdot \vec{n}_p)T^+(r)$$

where $\vec{n}_x, \vec{n}_p$ are the normals at x, P

For each node that descends, the contribution is simply the sum of its children's contributions. At the end, the root will return the overall diffuse reflectance at x. This is converted into radiance L_o

$$L_o(x, \vec{\omega}) = \frac{F_t(x, \vec{\omega})}{F_{dr}(x)} \frac{M_o(x)}{\pi}$$

where F_t is a Fresnel term, which can be omitted by assuming diffuse radiance.

5 Results

Our results show that our use of the empirical sum-of-Gaussian fit gives results of the same visual quality as using the direct multipole method. Additionally computing the combined profiles on a per-sample basis can be done about 30 times faster than with the FFT method. Below is a table showing relative timings for computing the overall profiles of a two-layered material. These timings were performed using an Intel E6550 Core 2 Duo with 4GB of RAM.

Table 1: Timings for Combining 2 Layers

FFT	.118 seconds
Our Approach	.004 seconds

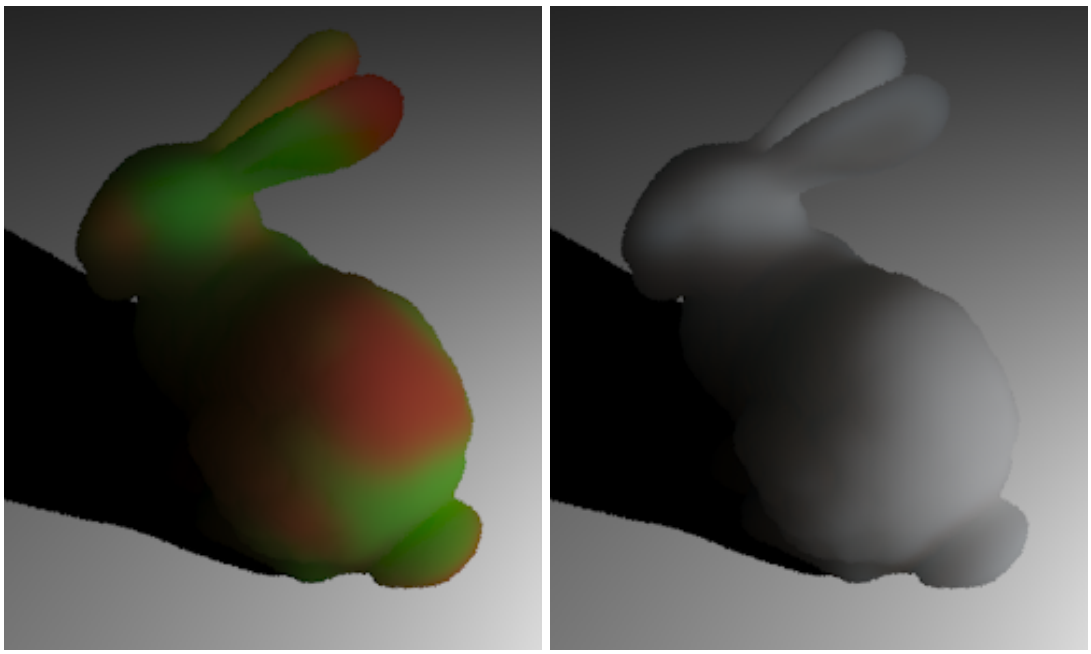
If we wanted to compute per-sample profiles, based on the timings above pre-computation time for a scene with 40,000 samples would take almost 5,000 seconds using the FFT method presented in Donner and Jensen (2005), while using our combined approach, it would only take about 160 seconds.

Figure 7 shows a picture of a multilayered bunny. The bottom layer is composed of a blend of two materials, one green, jade-like material, and a red, jasper-like material. The thickness of this layer is semi-infinite. On top of that sits a very thin (.78mm thick) layer of white marble. In the combined rendering, the color changes in the lower layer show up subtly through the top white layer.

Figure 8 shows a similar experiment, but now both layers have spatially changing parameters. Notice how in the combined rendering the detail change in the top layer is still visible, while at the same time the subtle contributions of the lower layer show up.

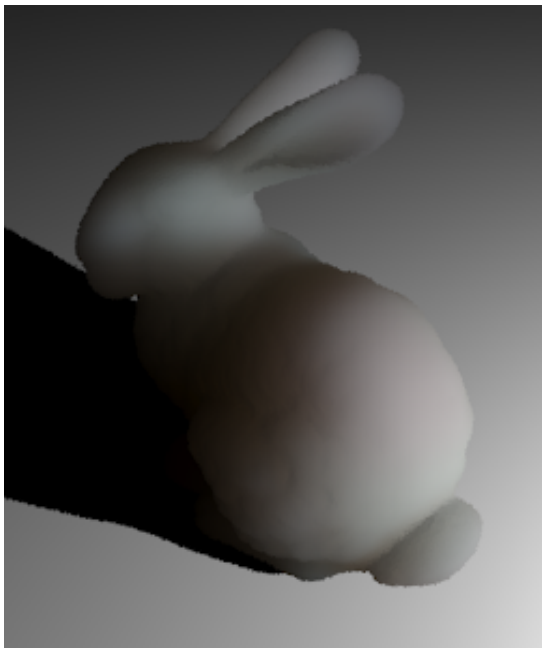
Figure 9 shows the types of effects that we can get by varying layer depth and scattering parameters on a per-sample basis. Here we simulate white veined marble with some imperfections. The thickness of the top layer was varied in the white regions of the bottom layer from 2mm-20mm, with 2mm being close to the mean free path. Since the assumptions that the multipole model makes break down when the thickness is much less then 2 times the mean free-path, we ignore the top layer if the thickness for a particular sample is less than 1.5 times the mean free-path and assume that light travels straight to the lower layer. While this is not entirely correct it still gives reasonable results.

Figure 10 shows the end result in full size.



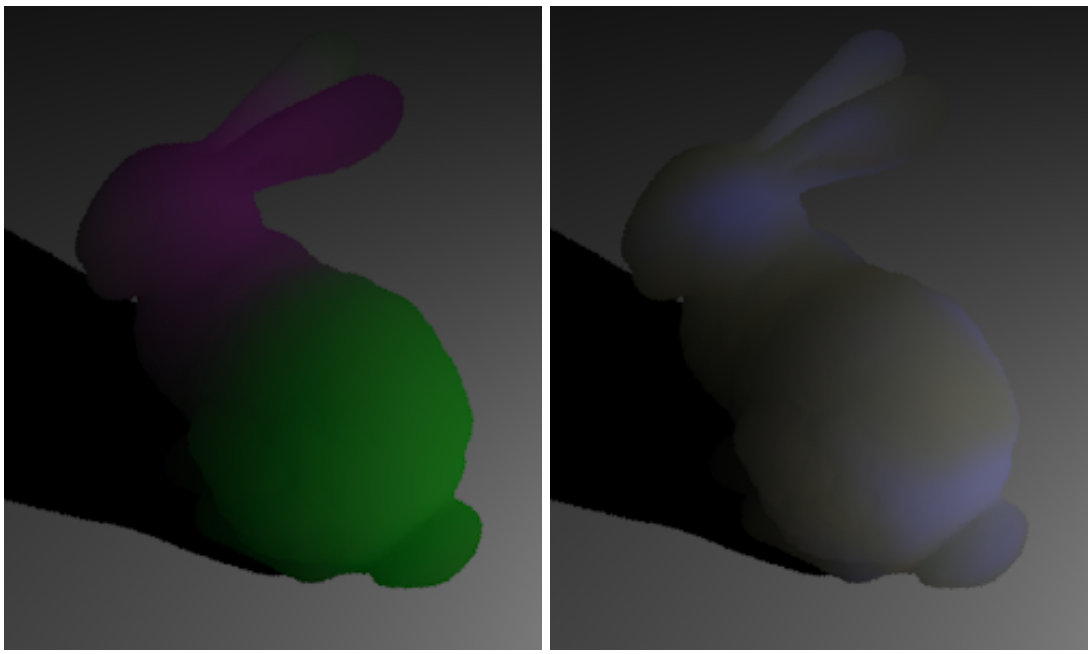
(a) Bottom layer

(b) Thin top layer



(c) Combined result

Figure 7: Here, a thick red-green bottom layer is combined with a very thin marble top layer

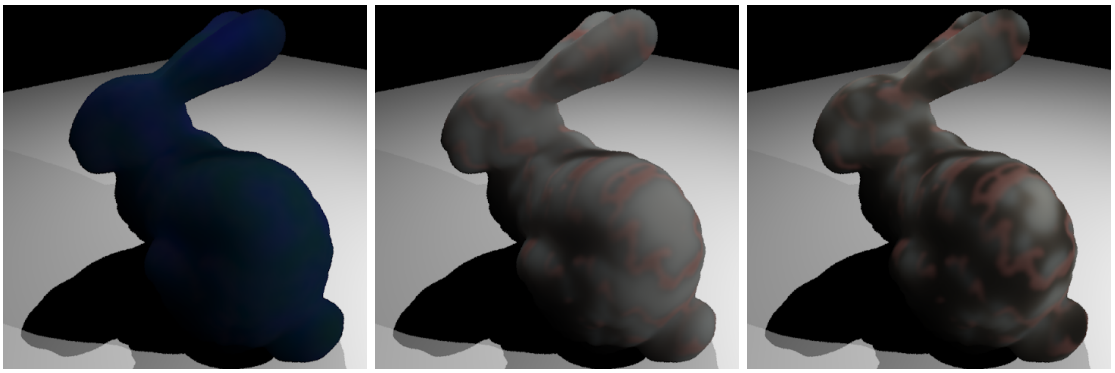


(a) Bottom layer

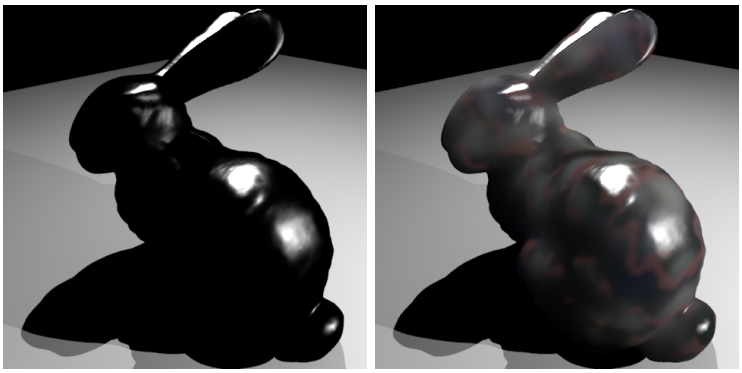
(b) Thin top layer

(c) Combined result

Figure 8: Here, a thick green-purple bottom layer is combined with a very top layer which alternates between highly scattering and lowly scattering. This picture illustrates that it is possible to vary both layers.



(a) Bottom layer (b) Top layer (c) Top layer with depth map applied



(d) Glossy reflectance (e) Combined result

Figure 9: Here we illustrate the types of effects that are now possible. The thickness of the top layer is varied from 2-20mm, shown in (c). Figure 10 shows (e) in original resolution.

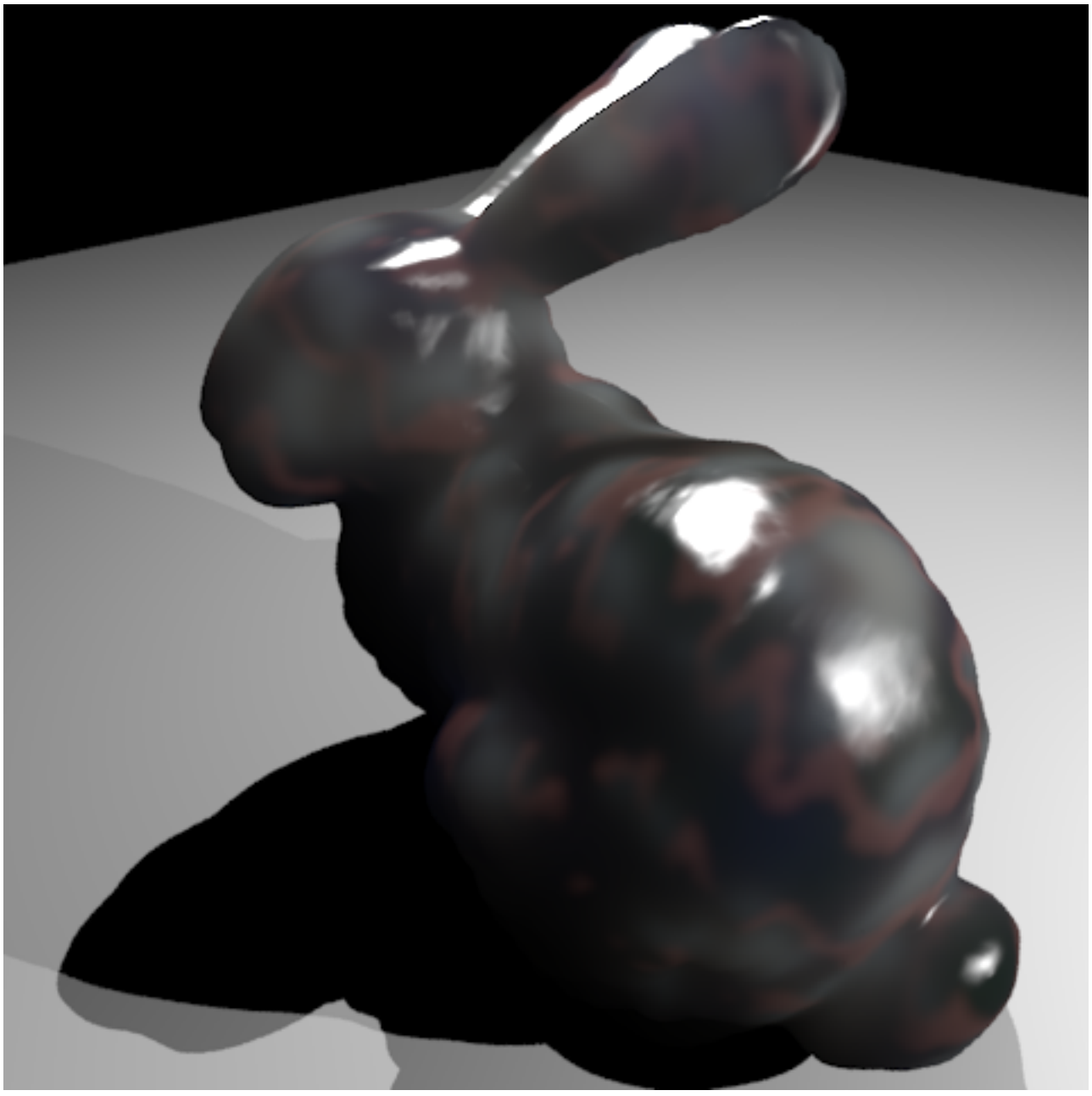


Figure 10: Full size picture of Figure 9 (e).

6 Conclusion and Future Work

We have introduced a novel approach to subsurface scattering rendering by combining two previous techniques. Our new approach allows us to vary scattering parameters using a texture map, with considerable speedup when compared to previous approaches. This allows us to render images with an added level of realism at a fraction of the time that it would take using previous methods.

Although we implemented our system in an off-line ray-tracer, it would be interesting to see how the approach maps to existing and future graphics hardware. There, the most important thing would be to see how to build and traverse an octree within the graphics card. Additionally, in our approach we only exploited one of the several advantages to using a sum-of-Gaussian diffusion profile approximation. Other properties, such as the computation of a wider Gaussian from a narrower one, could be exploited by using a specially tailored hierarchical integration method. Also, the system that we implemented has not been optimized, so there is much room for further improvements in speed.

7 Acknowledgments

Thanks to Eugene d'Eon for answering the many questions on his work. Also, thanks to Dr. Mark for introducing me to this project and to Dr. Fussell for the guidance through the last stages of this work. I would also like to thank all of the members of the Graphics group, my family, and friends for all of their continued support, critiques and ideas.

References

Subrahmanyan Chandrasekhar. *Radiative Transfer*. Dover Publications, New York, NY, USA, 1960.

Carsten Dachsbacher and Marc Stamminger. Translucent Shadow Maps . pages 197–201, Leuven, Belgium, 2003. Eurographics Association. ISBN 3-905673-03-7.

Eugene d'Eon, David Luebke, and Eric Enderton. Efficient Rendering of Human Skin. pages 147–157, Grenoble, France, 2007. Eurographics Association. ISBN 978-3-905673-52-4. doi: 10.2312/EGWR/EGSR07/147-157.

Craig Donner and Henrik Wann Jensen. Light diffusion in multi-layered translucent materials. *ACM Trans. Graph.*, 24(3):1032–1039, 2005. ISSN 0730-0301. doi: http://doi.acm.org/10.1145/1073204.1073308.

Julie Dorsey, Alan Edelman, Henrik Wann Jensen, Justin Legakis, and Hans Køhling Pedersen. Modeling and rendering of weathered stone. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 225–234, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co. ISBN 0-201-48560-5. doi: http://doi.acm.org/10.1145/311535.311560.

W. G. Egan, T. Hilgeman, and J. Reichman. Determination of absorption and scattering coefficients for nonhomogeneous media. 2: Experiment. *Applied Optics*, 12:1816–1832, 1973.

Pat Hanrahan and Wolfgang Krueger. Reflection from layered surfaces due to subsurface scattering. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 165–174, New York, NY, USA, 1993. ACM. ISBN 0-89791-601-8. doi: http://doi.acm.org/10.1145/166117.166139.

Henrik Wann Jensen. Rendering caustics on non-lambertian surfaces. In *GI '96: Proceedings of the conference on Graphics interface '96*, pages 116–121, Toronto, Ont., Canada, Canada, 1996. Canadian Information Processing Society. ISBN 0-9695338-5-3.

Henrik Wann Jensen and Juan Buhler. A rapid hierarchical rendering technique for translucent materials. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 576–581, New York, NY, USA, 2002. ACM. ISBN 1-58113-521-1. doi: http://doi.acm.org/10.1145/566570.566619.

Henrik Wann Jensen, Stephen R. Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 511–518, New York, NY, USA, 2001. ACM. ISBN 1-58113-374-X. doi: http://doi.acm.org/10.1145/383259.383319.

F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometrical considerations and nomenclature for reflectance. pages 94–145, 1992.

Matt Pharr and Pat Hanrahan. Monte carlo evaluation of non-linear scattering equations for subsurface reflection. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 75–84, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. ISBN 1-58113-208-5. doi: http://doi.acm.org/10.1145/344779.344824.

Matt Pharr and Greg Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004. ISBN 012553180X.